

University of Groningen

## Thermodynamics of the binary symmetric channel

Verbitskiy, Evgeny

*Published in:*  
Pacific journal of mathematics for industry

*DOI:*  
[10.1186/s40736-015-0021-5](https://doi.org/10.1186/s40736-015-0021-5)

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2016

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Verbitskiy, E. (2016). Thermodynamics of the binary symmetric channel. *Pacific journal of mathematics for industry*, 8, [2]. <https://doi.org/10.1186/s40736-015-0021-5>

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

ORIGINAL ARTICLE

Open Access



# Thermodynamics of the binary symmetric channel

Evgeny Verbitskiy<sup>1,2</sup>

## Abstract

We study a hidden Markov process which is the result of a transmission of the binary symmetric Markov source over the memoryless binary symmetric channel. This process has been studied extensively in Information Theory and is often used as a benchmark case for the so-called denoising algorithms. Exploiting the link between this process and the 1D Random Field Ising Model (RFIM), we are able to identify the Gibbs potential of the resulting Hidden Markov process. Moreover, we obtain a stronger bound on the memory decay rate. We conclude with a discussion on implications of our results for the development of denoising algorithms.

**Keywords:** Hidden Markov models, Gibbs states, Thermodynamic formalism, Denoising

**2000 Mathematics Subject Classification:** 37D35, 82B20, 82B20

## 1 Introduction

We study the binary symmetric Markov source over the memoryless binary symmetric channel. More specifically, let  $\{X_n\}$  be a stationary two-state Markov chain with values  $\{\pm 1\}$ , and

$$\mathbb{P}(X_{n+1} \neq X_n) = p,$$

where  $0 < p < 1$ ; denote by  $P = (p_{x,x'}) = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$  the corresponding transition probability matrix. Note that  $\pi = (\frac{1}{2}, \frac{1}{2})$  is the stationary initial distribution for this chain.

The binary symmetric channel will be modelled as a sequence of Bernoulli random variables  $\{Z_n\}$  with

$$\mathbb{P}_Z(Z_n = -1) = \varepsilon, \quad \mathbb{P}_Z(Z_n = 1) = 1 - \varepsilon, \quad \varepsilon \in (0, 1).$$

Finally, put

$$Y_n = X_n \cdot Z_n \quad (1.1)$$

for all  $n$ . The process  $\{Y_n\}$  is a Hidden Markov process, because  $Y_n \in \{-1, 1\}$  is chosen independently for any  $n$  from an *emission* distribution  $\pi_{X_n}$  on  $\{-1, 1\}$ :  $\pi_1 = (\varepsilon, 1 - \varepsilon)$  and  $\pi_{-1} = (1 - \varepsilon, \varepsilon)$ . More generally, the Hidden

Markov Processes are random functions of discrete-time Markov chains, where the value  $Y_n$  is chosen according to the distribution which depends on the value  $X_n = x_n$  of the underlying Markov chain, independently for any  $n$ . The applications of Hidden Markov processes include automatic character and speech recognition, information theory, statistics and bioinformatics, see [4, 13]. The particular example (1.1) we consider in the present paper is probably one of simplest examples, and often used as a benchmark for testing algorithms.

In particular, this example has been studied rather extensively in connection to computation of entropy of the output process  $\{Y_n\}$ , see e.g., [8–10, 12].

The law  $\mathbb{Q}$  of the process  $\{Y_n\}$  is the push-forward of  $\mathbb{P} \times \mathbb{P}_Z$  under  $\psi : \{-1, 1\}^{\mathbb{Z}} \times \{-1, 1\}^{\mathbb{Z}} \mapsto \{-1, 1\}^{\mathbb{Z}}$ , with  $\psi((x_n, z_n)) = x_n \cdot z_n$ . We write  $\mathbb{Q} = (\mathbb{P} \times \mathbb{P}_Z) \circ \psi^{-1}$ . For every  $m \leq n$ , and  $y_m^n := (y_m, \dots, y_n) \in \{-1, 1\}^{n-m+1}$ , the measure of the corresponding cylindric set is given by

$$\begin{aligned} \mathbb{Q}(y_m^n) &:= \mathbb{Q}(Y_m = y_m, \dots, Y_n = y_n) \\ &= \sum_{x_m^n, z_m^n \in \{-1, 1\}^{n-m+1}} \mathbb{P}(x_m^n) \mathbb{P}_Z(z_m^n) \prod_{k=m}^n \mathbb{I}[y_k = x_k \cdot z_k] \\ &= \sum_{x_m^n \in \{-1, 1\}^{n-m+1}} \frac{1}{2} \prod_{i=m}^{n-1} p_{x_i, x_{i+1}} \cdot \varepsilon^{\#\{i \in [m, n]: x_i y_i = -1\}} \\ &\quad \times (1 - \varepsilon)^{\#\{i \in [m, n]: x_i y_i = 1\}}. \end{aligned} \quad (1.2)$$

Correspondence: evgeny@math.leidenuniv.nl

<sup>1</sup>Mathematical Institute, Leiden University, Postbus 9512, 2300 RA Leiden, The Netherlands

<sup>2</sup>Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, PO Box 407, 9700 AK Groningen, The Netherlands

Two particular cases are easy to analyze. If  $p = \frac{1}{2}$ , then  $\{X_n\}$  is a sequence of independent identically distributed random variables with  $\mathbb{P}(X_n = -1) = \mathbb{P}(X_n = +1) = \frac{1}{2}$ , and  $\{Y_n\}$  has the same distribution. If  $\varepsilon = \frac{1}{2}$ , then the formula above implies that

$$\begin{aligned}\mathbb{Q}(y_m^n) &= \sum_{x_m^n \in \{-1,1\}^{n-m+1}} \frac{1}{2} \prod_{i=m}^{n-1} p_{x_i, x_{i+1}} \left(\frac{1}{2}\right)^{n-m+1} \\ &= \left(\frac{1}{2}\right)^{n-m+1},\end{aligned}$$

and hence again,  $\{Y_n\}$  is a sequence of independent random variables with  $\mathbb{Q}(Y_n = -1) = \mathbb{Q}(Y_n = +1) = \frac{1}{2}$ .

The paper is organized as follows. In Section 2 we exploit methods of Statistical Mechanics to derive expressions for probabilities of cylindric events (1.2). In Section 3, analyzing derived expressions, we show that the measure  $\mathbb{Q}$  has nice thermodynamic properties; in particular, it falls into the class of  $g$ -measures with exponential decay of variation (memory decay). We also obtain an novel estimate of the decay rate, which is stronger than estimates derived in previous works. In Section 4 we study two-sided conditional probabilities and show that  $\mathbb{Q}$  is in fact a Gibbs state in the sense of Statistical Mechanics. We also discuss well-known denoising algorithm DUDE, and suggest that the Gibbs property of  $\mathbb{Q}$  explains why DUDE performs so well in this particular example. Furthermore, we argue that the development of denoising algorithms, relying on thermodynamic Gibbs ideas can result in a superior performance.

## 2 Random field Ising model

It was observed in [18] that the probability  $\mathbb{Q}(y_m, \dots, y_n)$  of a cylindric event  $\{Y_m = y_m, \dots, Y_n = y_n\}$ ,  $m \leq n$ , can be expressed via a partition function of a random field Ising model. We exploit this observation further. Assume  $p, \varepsilon \in (0, 1)$ , and put

$$J = \frac{1}{2} \log \frac{1-p}{p}, \quad K = \frac{1}{2} \log \frac{1-\varepsilon}{\varepsilon}.$$

Then for any  $(y_m, \dots, y_n) \in \{-1, 1\}^{n-m+1}$ , expression for the cylinder probability (1.2) can be rewritten as

$$\begin{aligned}\mathbb{Q}(y_m, \dots, y_n) &= \frac{c_J}{\lambda_{J,K}^{n-m+1}} \sum_{x_m^n \in \{-1,1\}^{n-m+1}} \\ &\quad \exp \left( J \sum_{i=m}^{n-1} x_i x_{i+1} + K \sum_{i=m}^n x_i y_i \right),\end{aligned}$$

where

$$\begin{aligned}c_J &= \cosh(J), \quad \lambda_{J,K} = 2(\cosh(J+K) + \cosh(J-K)) \\ &= 4 \cosh(J) \cosh(K).\end{aligned}$$

The non-trivial part of the cylinder probability is the sum over all hidden configurations  $(x_m, \dots, x_n)$ :

$$Z_{m,n}(y_m^n) := \sum_{x_m^n \in \{-1,1\}^{n-m+1}} \exp \left( J \sum_{i=m}^{n-1} x_i x_{i+1} + K \sum_{i=m}^n x_i y_i \right)$$

is in fact the partition function of the Ising model with the signs of the external random fields given by  $y_i$ 's. Applying the recursive method of [14], the partition function can be evaluated in the following fashion [1]. Consider the following functions

$$\begin{aligned}A(w) &= \frac{1}{2} \log \frac{\cosh(w+J)}{\cosh(w-J)}, \\ B(w) &= \frac{1}{2} \log [4 \cdot \cosh(w+J) \cosh(w-J)] \\ &= \frac{1}{2} \log [e^{2w} + e^{-2w} + e^{2J} + e^{-2J}].\end{aligned}$$

One readily checks that if  $s = \pm 1$ , then for all  $w \in \mathbb{R}$

$$\exp(sA(w) + B(w)) = 2 \cosh(w + sJ). \quad (2.1)$$

Now the partition function can be evaluated by summing the right-most spin. Namely, suppose  $m < n$ ,  $y_m^n \in \{-1, 1\}^{n-m+1}$ , then

$$\begin{aligned}Z_{m,n}(y_m^n) &= \sum_{x_m^{n-1} \in \{-1,1\}^{m-n}} \exp \left( J \sum_{i=m}^{n-2} x_i x_{i+1} + K \sum_{i=m}^{n-1} x_i y_i \right) \\ &\quad \times \sum_{x_n \in \{-1,1\}} e^{x_n(Jx_{n-1} + Ky_n)} \\ &= \sum_{x_m^{n-1} \in \{-1,1\}^{m-n}} \exp \left( J \sum_{i=m}^{n-2} x_i x_{i+1} + K \sum_{i=m}^{n-1} x_i y_i \right) \\ &\quad \times \{2 \cosh(Jx_{n-1} + Ky_n)\} \\ &= \sum_{x_m^{n-1} \in \{-1,1\}^{m-n}} \exp \left( J \sum_{i=m}^{n-2} x_i x_{i+1} + K \sum_{i=m}^{n-1} x_i y_i \right) \\ &\quad \exp \left( x_{n-1} A(w_n^{(n)}) + B(w_n^{(n)}) \right)\end{aligned}$$

where

$$w_n^{(n)} = Ky_n.$$

Hence,

$$Z_{m,n}(y_m^n) = \sum_{x_m^{n-1} \in \{-1,1\}^{m-n}} \exp \left( J \sum_{i=m}^{n-2} x_i x_{i+1} + K \sum_{i=m}^{n-2} x_i y_i + x_{n-1} \left( \underbrace{Ky_{n-1} + A(w_n^{(n)})}_{w_{n-1}^{(n)}} \right) \right) \times \exp \left( B(w_n^{(n)}) \right),$$

and thus the new sum has exactly the same form, but instead of  $Ky_{n-1}$ , we now have  $w_{n-1}^{(n)} = Ky_{n-1} + A(w_n^{(n)})$ . Continuing the summation over the remaining right-most  $x$ -spins, one gets

$$Z_{m,n}(y_m^n) = 2 \cosh(w_m^{(n)}) \exp \left( \sum_{i=m+1}^n B(w_i^{(n)}) \right),$$

where

$$w_i^{(n)} = Ky_i + A(w_{i+1}^{(n)}) \quad \text{for every } i < n,$$

equivalently, since  $A(0) = 0$ , we can define

$$w_i^{(n)} = 0 \quad \forall i > n, \text{ and } w_i^{(n)} = Ky_i + A(w_{i+1}^{(n)}) \quad \forall i \leq n.$$

Therefore, we obtain the following expressions for the cylinder and conditional probabilities

$$\begin{aligned} \mathbb{Q}(y_0^n) &= \frac{c_J}{\lambda_{J,K}^{n+1}} \cosh(w_0^{(n)}) \exp \left( \sum_{i=1}^n B(w_i^{(n)}) \right), \\ \mathbb{Q}(y_0 | y_1^n) &= \frac{1}{\lambda_{J,K}} \frac{\cosh(w_0^{(n)}) \exp(B(w_1^{(n)}))}{\cosh(w_1^{(n)})}. \end{aligned} \quad (2.2)$$

### 3 Thermodynamic formalism

Let  $\Omega = A^{\mathbb{Z}_+}$ , where  $A$  is a finite alphabet, be the space of one-sided infinite sequences  $\omega = (\omega_0, \omega_1, \dots)$  in alphabet  $A$  ( $\omega_i \in A$  for all  $i$ ). We equip  $\Omega$  with the metric

$$d(\omega, \tilde{\omega}) = 2^{-k(\omega, \tilde{\omega})},$$

where  $k(\omega, \tilde{\omega}) = 1$  if  $\omega_0 \neq \tilde{\omega}_0$ , and  $k(\omega, \tilde{\omega}) = \max\{k \in \mathbb{N} : \omega_i = \tilde{\omega}_i \quad \forall i = 0, \dots, k-1\}$ , otherwise. Denote by  $S : \Omega \rightarrow \Omega$  the left shift:

$$(S\omega)_i = \omega_{i+1} \text{ for all } i \in \mathbb{Z}_+.$$

Borel probability measure  $\mathbb{P}$  is translation invariant if

$$\mathbb{P}(S^{-1}C) = \mathbb{P}(C)$$

for any Borel event  $C \subseteq \Omega$ .

Let us recall the following well-known definitions:

**Definition 3.1.** Suppose  $\mathbb{P}$  is a fully supported translation invariant measure on  $\Omega = A^{\mathbb{Z}_+}$ , where  $A$  is a finite alphabet.

(i) The measure  $\mathbb{P}$  is called a *g-measure*, if for some positive continuous function  $g : \Omega \rightarrow (0, 1)$  satisfying the normalization condition

$$\sum_{\tilde{\omega}_0 \in A} g(\tilde{\omega}_0, \omega_1, \omega_2, \dots) = 1$$

for all  $\omega = (\omega_0, \omega_1, \dots) \in \Omega$ , one has

$$\mathbb{P}(\omega_0 | \omega_1, \omega_2, \dots) = g(\omega_0, \omega_1, \dots)$$

for  $\mathbb{P}$ -a.a.  $\omega \in \Omega$ .

(ii) The measure  $\mathbb{P}$  is *Bowen-Gibbs* for a continuous potential  $\phi : \Omega \rightarrow \mathbb{R}$ , if there exist constants  $P \in \mathbb{R}$  and  $C \geq 1$  such that for all  $\omega \in \Omega$  and every  $n \in \mathbb{N}$

$$\frac{1}{C} \leq \frac{\mathbb{P}(\{\tilde{\omega} \in \Omega : \tilde{\omega}_0 = \omega_0, \dots, \tilde{\omega}_{n-1} = \omega_{n-1}\})}{\exp((S_n \phi)(\omega) - nP)} \leq C,$$

where  $(S_n \phi)(\omega) = \sum_{k=0}^{n-1} \phi(S^k \omega)$ .

(iii) The measure  $\mathbb{P}$  is called an *equilibrium state* for continuous potential  $\phi : \Omega \rightarrow \mathbb{R}$ , if  $\mathbb{P}$  attains maximum of the following functional

$$h(\mathbb{P}) + \int \phi d\mathbb{P} = \sup_{\tilde{\mathbb{P}} \in \mathcal{M}_1^*(\Omega)} \left[ h(\tilde{\mathbb{P}}) + \int \phi d\tilde{\mathbb{P}} \right], \quad (3.1)$$

where  $h(\mathbb{P})$  is the Kolmogorov-Sinai entropy of  $\mathbb{P}$  and the supremum is taken over the set  $\mathcal{M}_1^*(\Omega)$  of all translation invariant Borel probability measures on  $\Omega$ .

It is known that every  $g$ -measure  $\mathbb{P}$  is also an equilibrium state for  $\log g$ ; and that every Bowen-Gibbs measure  $\mathbb{P}$  for potential  $\phi$  is an equilibrium state for  $\phi$  as well.

**Theorem 3.1.** Suppose  $p, \varepsilon \in (0, 1)$ . Then the measure  $\mathbb{Q} = \mathbb{Q}_{p,\varepsilon}$  on  $\{-1, 1\}^{\mathbb{Z}_+}$  (c.f., (2.2)) is a  $g$ -measure. Moreover, the corresponding function  $g$  has exponential decay of variation: define the  $n$ -th variation of  $g$  by

$$\text{var}_n(g) := \sup_{y, \tilde{y} : y_0^{n-1} = \tilde{y}_0^{n-1}} |g(y) - g(\tilde{y})|, \quad n \geq 1,$$

then

$$\rho(p, \varepsilon) = \limsup_{n \rightarrow \infty} (\text{var}_n(g))^{\frac{1}{n}} < 1. \quad (3.2)$$

Furthermore,  $\rho(p, \varepsilon) = 0$  if  $p = \frac{1}{2}$  or  $\varepsilon = \frac{1}{2}$ ; for all  $p \neq \frac{1}{2}$

$$\rho(p, \varepsilon) < |1 - 2p|. \quad (3.3)$$

Finally, the measure  $\mathbb{Q}$  is also a Bowen-Gibbs measure for a Hölder continuous potential  $\phi : \{-1, 1\}^{\mathbb{Z}_+} \rightarrow \mathbb{R}$ .

The result of Theorem 3.1 is actually true in much greater generality: namely, for distributions of functions of Markov chains  $\{Y_n\}$ , where the underlying Markov chain  $\{X_n\}$  has strictly positive transition probability matrix  $P$ ,

see [15] for review of several results of this nature. However, example considered in the present paper is rather exceptional since one is able to identify the  $g$ -function and the Gibbs potential  $\phi$  *explicitly*. Another interesting question is the estimate of the decay rate  $\rho$ . In [15] a number of previously known estimates of the rate of exponential decay in (3.3) have been compared; the best known estimate for  $\rho$

$$\rho \leq |1 - 2p|$$

is due to [8] and [7]. Quite surprisingly this estimate does not depend on  $\varepsilon$ , and in fact, it was conjectured in [15] that the estimate could be improved, e.g., by incorporating dependence on  $\varepsilon$ . The proof of Theorem 3.1 shows that this is indeed the case and one obtains sharper estimate (3.3).

Let us start by introducing some notation and proving a technical result. Suppose  $p, \varepsilon \in (0, 1)$ , and  $p \neq \frac{1}{2}$  and  $\varepsilon \neq \frac{1}{2}$ . Fix  $\mathbf{y} = (y_0, y_1, \dots) \in \{-1, 1\}^{\mathbb{Z}_+}$ . For every  $n \in \mathbb{Z}_+$ , define the sequence  $w_i^{(n)} = w_i^{(n)}(\mathbf{y})$ ,  $i \in \mathbb{Z}_+$ , as follows:

$$\begin{aligned} w_i^{(n)} &= 0, \text{ for every } i \geq n+1, \\ w_i^{(n)} &= Ky_i + A(w_{i+1}^{(n)}), \text{ for } i \leq n. \end{aligned}$$

If we introduce maps  $F_{-1}, F_1 : \mathbb{R} \rightarrow \mathbb{R}$ , given by

$$F_{-1}(w) = -K + A(w), \quad F_1(w) = K + A(w),$$

then for  $i \leq n$ ,

$$\begin{aligned} w_i^{(n)} &= w_i^{(n)}(\mathbf{y}) = F_{y_i}(F_{y_{i+1}}(\dots(F_{y_n}(0))\dots)) \\ &= F_{y_i} \circ F_{y_{i+1}} \circ \dots \circ F_{y_n}(0). \end{aligned} \quad (3.4)$$

As we will show the maps  $F_{-1}, F_1$  are strict contractions, and as a result, for every  $i$ , the sequence  $\{w_i^{(n)}\}$  is converging as  $n \rightarrow \infty$ ; in fact, with exponential speed.

**Lemma 3.2.** *For every  $i \in \mathbb{Z}_+$  and all  $\mathbf{y} \in \{-1, 1\}^{\mathbb{Z}_+}$  one has*

$$\lim_{n \rightarrow \infty} w_i^{(n)}(\mathbf{y}) =: w_i(\mathbf{y}).$$

Moreover, there exist constants  $\varrho \in (0, 1)$  and  $C > 0$ , both independent of  $\mathbf{y}$ , such that

$$|w_i^{(n)}(\mathbf{y}) - w_i(\mathbf{y})| \leq C\varrho^{n-i} \quad (3.5)$$

for all  $n \geq i$ . Furthermore,  $w_i(\mathbf{y}) = w_0(S^i \mathbf{y})$  for all  $i \in \mathbb{Z}_+$  and  $\mathbf{y}$ , and  $w_0 : \{-1, 1\}^{\mathbb{Z}_+} \rightarrow \mathbb{R}$  (and hence every  $w_i$ ) is Hölder continuous

$$|w_0(\mathbf{y}) - w_0(\tilde{\mathbf{y}})| \leq C' (d(\mathbf{y}, \tilde{\mathbf{y}}))^\theta$$

for some  $C', \theta > 0$  and all  $\mathbf{y}, \tilde{\mathbf{y}} \in \{-1, 1\}^{\mathbb{Z}_+}$ .

*Proof.* Suppose  $i \leq n \leq m$ . Then

$$\begin{aligned} |w_i^{(n)} - w_i^{(m)}| &= |A(w_{i+1}^{(n)}) - A(w_{i+1}^{(m)})| \\ &\leq |w_{i+1}^{(n)} - w_{i+1}^{(m)}| \cdot \sup_w \left| \frac{dA}{dw} \right|. \end{aligned}$$

One has

$$\frac{dA}{dw} = \frac{\sinh(2J)}{\cosh(2J) + \cosh(2w)},$$

and hence

$$\varrho := \sup_w \left| \frac{dA}{dw} \right| = \left| \frac{\sinh(2J)}{\cosh(2J) + 1} \right| = |1 - 2p| < 1. \quad (3.6)$$

Combined with the fact that for all  $i \in \mathbb{Z}_+$

$$\begin{aligned} |w_i^{(m)}| &= |Ky_i + A(w_{i+1}^{(m)})| \leq |K| + |\operatorname{arctanh}(1 - 2p)| \\ &\leq |K| + |J| =: C_1, \end{aligned}$$

one therefore concludes that for  $i \leq n \leq m$

$$\begin{aligned} |w_i^{(n)} - w_i^{(m)}| &\leq \varrho^{n-i+1} |w_{n+1}^{(n)} - w_{n+1}^{(m)}| = \varrho^{n-i+1} |w_{n+1}^{(m)}| \\ &\leq C_1 \varrho^{n-i+1}. \end{aligned}$$

Hence,  $\lim_{n \rightarrow \infty} w_i^{(n)} =: w_i$  exists and

$$\begin{aligned} |w_i^{(n)} - w_i| &\leq \sum_{m=n}^{\infty} |w_i^{(m)} - w_i^{(m+1)}| \leq C_1 \sum_{m=n}^{\infty} \varrho^{m-i+1} \\ &= \frac{C_1}{1 - \varrho} \varrho^{n-i+1} =: C \varrho^{n-i+1}. \end{aligned}$$

From representation (3.4) it is clear that for  $n \geq i$ ,

$$w_i^{(n)}(\mathbf{y}) = w_0^{(n-i)}(S^i \mathbf{y}),$$

and hence  $w_i(\mathbf{y}) = w_0(S^i \mathbf{y})$ .

Suppose  $\mathbf{y} = (y_0, y_1, \dots), \tilde{\mathbf{y}} = (\tilde{y}_0, \tilde{y}_1, \dots) \in \{-1, 1\}^{\mathbb{Z}_+}$  are such that  $d(\mathbf{y}, \tilde{\mathbf{y}}) = 2^{-k}$  for some  $k \in \mathbb{N}$ , i.e.,  $y_0 = \tilde{y}_0, \dots, y_{k-1} = \tilde{y}_{k-1}$ . Then

$$\begin{aligned} |w_0(\mathbf{y}) - w_0(\tilde{\mathbf{y}})| &= |F_{y_0} \circ \dots \circ F_{y_{k-1}}(w_k(\mathbf{y})) - F_{y_0} \circ \dots \circ F_{y_{k-1}}(w_k(\tilde{\mathbf{y}}))| \\ &\leq \sup_{w \in \mathbb{R}} |(F_{y_0} \circ \dots \circ F_{y_{k-1}})'(w)| \cdot |w_k(\mathbf{y}) - w_k(\tilde{\mathbf{y}})| \\ &\leq \left( \sup_{w \in \mathbb{R}} |A'(w)| \right)^k \cdot (2C_1) = 2C_1 \varrho^k, \end{aligned}$$

and hence  $w_0 : \{-1, 1\}^{\mathbb{Z}_+} \rightarrow \mathbb{R}$  is Hölder continuous.  $\square$

The estimate of a contraction rate in the Lemma above can be improved. If  $p = \frac{1}{2}$ , then  $A(w) \equiv 0$ , and hence  $w_i^{(n)} \equiv Ky_i$  for all  $n \geq i$ . We may assume that  $p \neq \frac{1}{2}$ . Let us also assume that  $\varepsilon \neq \frac{1}{2}$ , i.e.,  $K \neq 0$ .

Let us now consider second iterations:

$$\begin{aligned} |w_i^{(n)} - w_i^{(m)}| &= |A(w_{i+1}^{(n)}) - A(w_{i+1}^{(m)})| = |A(Ky_{i+1} + A(w_{i+2}^{(n)})) \\ &\quad - A(Ky_{i+1} + A(w_{i+2}^{(m)}))| \\ &\leq \left( \sup_w |A'(K + A(w))A'(w)| \right) |w_{i+2}^{(n)} - w_{i+2}^{(m)}| \\ &=: \rho^{(2)} |w_{i+2}^{(n)} - w_{i+2}^{(m)}|. \end{aligned}$$

We are going to show that for  $p \neq \frac{1}{2}$  and all  $\varepsilon \neq \frac{1}{2}$ , one has

$$\rho^{(2)} = \sup_w |A'(K + A(w))A'(w)| < (1 - 2p)^2. \quad (3.7)$$

Firstly, note that

$$\begin{aligned} &|A'(K + A(w))A'(w)| \\ &= \frac{\sinh^2(2J)}{(\cosh(2J) + \cosh(2w))(\cosh(2J) + \cosh(2K + 2A(w)))} \\ &= \frac{(1 - 2p)^2}{(\alpha + (1 - \alpha) \cosh(2K + 2A(w))) \cdot (\alpha + (1 - \alpha) \cosh(2w))}, \end{aligned} \quad (3.8)$$

where  $\alpha = (1 - p)^2 + p^2$ ,  $1 - \alpha = 2p(1 - p)$ . Let  $\Delta > 0$  be sufficiently small so that for all  $w \in [-\Delta, \Delta]$  one has  $\cosh(2K + 2A(w)) > \cosh(K)$ , and hence for all  $w \in [-\Delta, \Delta]$

$$|A'(K + A(w))A'(w)| \leq \frac{(1 - 2p)^2}{(\alpha + (1 - \alpha) \cosh(K)) \cdot 1} < (1 - 2p)^2.$$

For  $w \notin [-\Delta, \Delta]$ , one has

$$\begin{aligned} |A'(K + A(w))A'(w)| &\leq \frac{(1 - 2p)^2}{1 \cdot (\alpha + (1 - \alpha) \cosh(\Delta))} \\ &< (1 - 2p)^2. \end{aligned}$$

Hence,

$$\rho^{(2)} = \min \left\{ \frac{(1 - 2p)^2}{(\alpha + (1 - \alpha) \cosh(K))}, \frac{(1 - 2p)^2}{(\alpha + (1 - \alpha) \cosh(\Delta))} \right\} < (1 - 2p)^2,$$

and thus (3.5) holds for  $\bar{\rho} = \sqrt{\rho^{(2)}} < |1 - 2p|$  and some constant  $\tilde{C} > 0$ . In particular, we are now able conclude that

$$\begin{aligned} |w_0(\mathbf{y}) - w_0(\tilde{\mathbf{y}})| &\leq C_2 \bar{\rho}^{k(\mathbf{y}, \tilde{\mathbf{y}})} = C_2 (d(\mathbf{y}, \tilde{\mathbf{y}}))^\theta, \\ \theta &= -\log_2 \bar{\rho} > 0. \end{aligned} \quad (3.9)$$

Even sharper bounds can be achieved by studying minimum of the denominator in (3.8) or higher iterates of  $F$ 's.

*Proof of Theorem 3.1.* The cases  $p = \frac{1}{2}$  or  $\varepsilon = \frac{1}{2}$  are obvious: in these cases,  $\mathbb{Q}$  is the Bernoulli  $(\frac{1}{2}, \frac{1}{2})$ -measure on  $\{-1, 1\}^{\mathbb{Z}_+}$ , and hence  $\rho(p, \varepsilon) = 0$ . Thus we will assume that  $p, \varepsilon \neq \frac{1}{2}$ .

To show that  $\mathbb{Q}$  is a  $g$ -measure it is sufficient to show that conditional probabilities  $\mathbb{Q}(y_0|y_1^n)$  converge uniformly as  $n \rightarrow \infty$ . Given that

$$\mathbb{Q}(y_0|y_1^n) = \frac{1}{\lambda_{J,K}} \frac{\cosh(w_0^{(n)}) \exp(B(w_1^{(n)}))}{\cosh(w_1^{(n)})}, \quad (3.10)$$

and using the result of Lemma 3.2:  $w_i^{(n)}(\mathbf{y}) \Rightarrow w_i(\mathbf{y})$  as  $n \rightarrow \infty$ , we obtain uniform convergence of conditional probabilities, and hence,  $\mathbb{Q}$  is a  $g$ -measure with  $g$  given by

$$g(\mathbf{y}) = \frac{1}{\lambda_{J,K}} \frac{\cosh(w_0(\mathbf{y})) \exp(B(w_1(\mathbf{y})))}{\cosh(w_1(\mathbf{y}))}. \quad (3.11)$$

Taking into account that  $w_0, w_1 = w_0 \circ S$  are Hölder continuous functions satisfying (3.9), and that  $\cosh$ ,  $\exp$ , and  $B$  are smooth functions, we can conclude that  $g$  is also Hölder continuous with the same decay of variation:

$$\begin{aligned} \text{var}_n(g) &= \sup_{\mathbf{y}, \tilde{\mathbf{y}}: y_0^{n-1} = \tilde{y}_0^{n-1}} |g(\mathbf{y}) - g(\tilde{\mathbf{y}})| \\ &\leq C_3 |w_1(\mathbf{y}) - w_1(\tilde{\mathbf{y}})| \leq C_4 \bar{\rho}^{n-1}, \end{aligned}$$

for some  $C_3 > 0$  ( $C_4 = C_2 C_3$ , c.f., (3.9)), and hence

$$\rho(p, \varepsilon) = \limsup_{n \rightarrow \infty} (\text{var}_n(g))^{\frac{1}{n}} \leq \bar{\rho} < |1 - 2p|.$$

Let us introduce the following functions: for  $\mathbf{y} \in \{-1, 1\}^{\mathbb{Z}_+}$ , put

$$\phi(\mathbf{y}) = B(w_0(\mathbf{y})), \quad h(\mathbf{y}) = \cosh(w_0(\mathbf{y})) \exp(-B(w_0(\mathbf{y}))).$$

Taking into account that  $w_1(\mathbf{y}) = w_0(S\mathbf{y})$ , one has

$$g(\mathbf{y}) = \frac{e^{\phi(\mathbf{y})}}{\lambda_{J,K}} \frac{h(\mathbf{y})}{h(S\mathbf{y})}.$$

Since every  $g$ -measure is also an equilibrium state for  $\log g$ , we conclude that  $\mathbb{Q}$  is an equilibrium state for

$$\tilde{\phi}(\mathbf{y}) = \phi(\mathbf{y}) + \log h(\mathbf{y}) - \log h(S\mathbf{y}) - \log \lambda_{J,K}.$$

The difference  $\tilde{\phi}(\mathbf{y}) - \phi(\mathbf{y})$  has a very special form: it is a sum of a so-called coboundary ( $\log h(\mathbf{y}) - \log h(S\mathbf{y})$ ) and a constant ( $-\log \lambda_{J,K}$ ). Two potentials, whose difference is of a such form, have identical sets of equilibrium states. The reason is that for any translation invariant measure  $\mathbb{Q}'$  one has

$$\begin{aligned} \int (\log h(\mathbf{y}) - \log h(S\mathbf{y}) - \log \lambda_{J,K}) d\mathbb{Q}' &= -\log \lambda_{J,K} \\ &= \text{const.} \end{aligned}$$

Therefore, if  $\mathbb{Q}'$  achieves maximum in the righthand side of (3.1) for  $\tilde{\phi}$ , then  $\mathbb{Q}'$  achieves maximum for  $\phi$  as well. Thus  $\mathbb{Q}$  is also an equilibrium state for

$$\phi(y) = B(w_0(y)) = \frac{1}{2} \log \left[ 4 \sinh^2(w_0(y)) + \frac{1}{p(1-p)} \right].$$

Any equilibrium measure for a Hölder continuous potential  $\phi$  is also a Bowen-Gibbs measure [3]. In our particular case, direct proof of the Bowen-Gibbs property for  $\mathbb{Q}$  is straightforward. Indeed, using the result of (2.2) and the notation introduced above, for every  $y = (y_0, y_1, \dots)$  one has

$$\begin{aligned} \mathbb{Q}(y_0^n) &= \frac{c_J}{\lambda_{J,K}^{n+1}} \exp \left( \sum_{i=1}^n B(w_i^{(n)}(y)) \right) \cosh(w_0^{(n)}(y)) \\ &= \frac{c_J \cdot \cosh(w_0^{(n)}(y))}{\exp(B(w_0(y)))} \exp \left( \sum_{i=1}^n [B(w_i^{(n)}(y)) - B(w_i(y))] \right) \\ &\quad \times \exp \left( \sum_{i=0}^n B(w_i(y)) - (n+1) \log \lambda_{J,K} \right). \end{aligned}$$

Therefore, for  $P = \log \lambda_{J,K}$ ,

$$\frac{\mathbb{Q}(y_0^n)}{\exp((S_{n+1}\phi)(y) - (n+1)P)} = \frac{c_J \cdot \cosh(w_0^{(n)}(y))}{\exp(B(w_0(y)))} \times \exp \left( \sum_{i=1}^n [B(w_i^{(n)}(y)) - B(w_i(y))] \right)$$

It only remains to demonstrate that the right hand side is uniformly bounded (both in  $n$  and  $y = (y_0, y_1, \dots)$ ) from below and above by some positive constants  $\underline{C}, \bar{C}$ , respectively. Indeed, since  $p, \varepsilon \in (0, 1)$ ,  $I = [-|K| - |J|, |K| + |J|]$  is a finite interval, by the result of the previous Lemma,  $w_i^{(n)}(y) \in I$  for all  $i$  and  $n$ . Using (3.5), one readily checks that the following choice of constants suffices:

$$\begin{aligned} \bar{C} &= c_J \frac{\sup_{w \in I} \cosh(w)}{\inf_{w \in I} \exp(B(w))} \exp \left( \frac{C}{1-\varrho} \sup_{w \in I} \left| \frac{dB}{dw} \right| \right) < \infty, \\ \underline{C} &= c_J \frac{\inf_{w \in I} \cosh(w)}{\sup_{w \in I} \exp(B(w))} \exp \left( -\frac{C}{1-\varrho} \sup_{w \in I} \left| \frac{dB}{dw} \right| \right) > 0. \end{aligned}$$

□

We complete this section with a curious continued fraction representation of the  $g$ -function (3.11).

**Proposition 3.3.** For every  $y = (y_0, y_1, \dots) \in \{-1, 1\}^{\mathbb{Z}_+}$ , one has

$$2g(y) = a_1 - \frac{b_1}{a_2 - \frac{b_2}{a_3 - \frac{b_3}{a_4 - \dots}}} \quad (3.12)$$

where for  $i \geq 1$

$$q_i = (1-2p)y_{i-1}y_i, \quad a_i = 1 + q_i, \quad b_i = 4\varepsilon(1-\varepsilon)q_i.$$

*Proof.* Using elementary transformations, one can show that for every  $y = (y_0, y_1, \dots) \in \{-1, 1\}^{\mathbb{Z}_+}$  one has

$$\begin{aligned} g(y) &= \frac{1}{\lambda_{J,K}} \frac{\cosh(w_0(y))}{\cosh(w_1(y))} \exp(B(w_1(y))) \\ &= \frac{1}{2} + \frac{1}{2}(1-2p)(1-2\varepsilon)y_0 \tanh(w_1(y)). \end{aligned} \quad (3.13)$$

Since for all  $w \in \mathbb{R}$

$$\tanh(A(w)) = \tanh(J) \tanh(w) = (1-2p) \tanh(w),$$

for every  $i \in \mathbb{Z}_+$ , one has

$$\begin{aligned} \tanh(w_i) &= \frac{\tanh(Ky_i) + \tanh(A(w_{i+1}))}{1 + \tanh(Ky_i) \cdot \tanh(A(w_{i+1}))} \\ &= \frac{(1-2\varepsilon)y_i + (1-2p) \tanh(w_{i+1})}{1 + (1-2\varepsilon)(1-2p)y_i \tanh(w_{i+1})} \\ &= y_i \frac{(1-2\varepsilon) + (1-2p)y_i \tanh(w_{i+1})}{1 + (1-2\varepsilon)(1-2p)y_i \tanh(w_{i+1})}. \end{aligned}$$

Therefore, if we let  $z_i = (1-2p)(1-2\varepsilon)y_{i-1} \tanh(w_i)$ ,  $i \in \mathbb{N}$ , then

$$\begin{aligned} z_i &= (1-2p)y_{i-1}y_i - \frac{4\varepsilon(1-\varepsilon)(1-2p)y_{i-1}y_i}{1+z_{i+1}} \\ &= q_i - \frac{b_i}{1+z_{i+1}}. \end{aligned}$$

Since  $g(y) = \frac{1}{2} + \frac{1}{2}z_1$ , we obtain the continued fraction expansion (3.12). □

## 4 Two-sided conditional probabilities and denoising

In the previous section we established that  $\mathbb{Q}$  is a Bowen-Gibbs measure. The notion of a Gibbs measure originates in Statistical Mechanics, and is not equivalent to the Bowen-Gibbs definition. In Statistical Mechanics, one is interested in **two-sided** conditional probabilities

$$\mathbb{Q}(y_0|y_{-m}^{-1}, y_1^n) \text{ or } \mathbb{Q}(y_0|y_{<0}, y_{>0}) := \mathbb{Q}(y_0|y_{-\infty}^{-1}, y_1^{\infty}).$$

The method of Section 2 can be used to evaluate conditional probabilities  $\mathbb{Q}(y_0|y_{-m}^{-1}, y_1^n)$ ,  $m, n > 0$  for  $y = (\dots, y_{-1}, y_0, y_1, \dots) \in \{-1, 1\}^{\mathbb{Z}}$ . Indeed,

$$\mathbb{Q}(y_0|y_{-m}^{-1}, y_1^n) = \frac{\mathbb{Q}(y_{-m}^{-1}, y_0, y_1^n)}{\mathbb{Q}(y_{-m}^{-1}, y_0, y_1^n) + \mathbb{Q}(y_{-m}^{-1}, \bar{y}_0, y_1^n)},$$

where  $\bar{y}_0 = -y_0$ . We can evaluate

$$\begin{aligned} & \mathbb{Q}(y_{-m}, \dots, y_{-1}, y_0, y_1, \dots, y_n) \\ &= \frac{c_J}{\lambda_{J,K}^{n+m+1}} \sum_{x_{-m}^n \in \{-1, 1\}^{n+m+1}} \exp\left(J \sum_{i=-m}^{n-1} x_i x_{i+1} + K \sum_{i=-m}^n x_i y_i\right) \\ &= \frac{c_J}{\lambda_{J,K}^{n+m+1}} Z_{-m,n}(y_{-m}^n), \end{aligned}$$

by first summing over spins on the right:  $x_n, \dots, x_1$ , and then summing over spins on the left:  $x_{-m}, \dots, x_{-1}$ . One has

$$\begin{aligned} Z_{-m,n}(y_{-m}^n) &= \sum_{x_{-m}, \dots, x_0} \exp\left(J \sum_{i=-m}^{-1} x_i x_{i+1} + K \sum_{i=-m}^0 x_i y_i + x_0 A(w_1^{(n)})\right) \\ &\quad \exp\left(\sum_{i=1}^n B(w_i^{(n)})\right) \\ &= \exp\left(\sum_{j=-m}^{-1} B(w_j^{(-m)})\right) 2 \cosh(w_0^{(-m,n)}) \\ &\quad \exp\left(\sum_{i=1}^n B(w_i^{(n)})\right) \end{aligned}$$

where now  $w_{-m}^{(-m)} = Ky_{-m}$ ,

$$w_{j+1}^{(-m)} = Ky_{j+1} + A(w_j^{(-m)}), \quad j = -m, \dots, -2,$$

and

$$w_0^{(-m,n)} = Ky_0 + A(w_{-1}^{(-m)}) + A(w_1^{(n)}).$$

Therefore,

$$\begin{aligned} \mathbb{Q}(y_0|y_{-m}^{-1}, y_1^n) &= \frac{Z_{-m,n}(y_{-m}^{-1}, y_0, y_1^n)}{Z_{-m,n}(y_{-m}^{-1}, y_0, y_1^n) + Z_{-m,n}(y_{-m}^{-1}, \bar{y}_0, y_1^n)} \\ &= \frac{\cosh(Ky_0 + A(w_{-1}^{(-m)}) + A(w_1^{(n)}))}{\cosh(Ky_0 + A(w_{-1}^{(-m)}) + A(w_1^{(n)})) + \cosh(-Ky_0 + A(w_{-1}^{(-m)}) + A(w_1^{(n)}))}. \end{aligned}$$

Again, given this expression, one easily establishes uniform convergence and existence of the limits,

$$\mathbb{Q}(y_0|y_{-\infty}^{-1}, y_1^\infty) = \lim_{m,n \rightarrow \infty} \mathbb{Q}(y_0|y_{-m}^{-1}, y_1^n).$$

Thus the two sided conditional probabilities are also regular, c.f. Theorem 3.1.

#### 4.1 Denoising

Reconstruction of signals corrupted by noise during the transmission is one of the classical problems in Information Theory. Suppose we observe a sequence  $\{y_n\}$ ,  $n = 1, \dots, N$ , given by (1.1), i.e.,

$$y_n = x_n \cdot z_n.$$

where  $\{x_n\}$  is some unknown realisation of the Markov chain, and  $\{z_n\}$  is unknown realisation of the Bernoulli sequence  $\{Z_n\}$ . The natural question is, given the observed data  $y^N = (y_1, \dots, y_N)$ , what is the optimal choice of  $\hat{X}_n = \hat{X}_n(y^N)$  – the estimate of  $X_n$ , such that the empirical zero-one loss (bit error rate)

$$L_N = \frac{1}{N} \sum_{n=1}^N \mathbb{I}[\hat{X}_n \neq x_n]$$

is minimal. The corresponding standard maximum a posteriori probability (MAP) estimator (denoiser) is given by

$$\begin{aligned} \hat{X}^n &= \hat{X}^n(y^N) \\ &= \begin{cases} -1, & \text{if } \mathbb{P}[X_n = -1 | Y^N = y_1^N] \geq \mathbb{P}[X_n = 1 | Y^N = y_1^N], \\ +1, & \text{if } \mathbb{P}[X_n = -1 | Y^N = y_1^N] < \mathbb{P}[X_n = 1 | Y^N = y_1^N], \end{cases} \\ & n = 1, \dots, N. \end{aligned}$$

In case, parameters of the Markov chain (i.e.,  $P$ ) and of the channel (i.e.,  $\Pi$ ) are known, conditional probabilities  $\mathbb{P}[X_n = x | Y^N = y^N]$  can be found using the **backward-forward algorithm**. Namely, one has

$$\mathbb{P}[X_n = x | Y^N = y^N] = \frac{\alpha_n(x) \beta_n(x)}{\sum_{\tilde{x} \in \mathcal{A}} \alpha_n(\tilde{x}) \beta_n(\tilde{x})} \quad (4.1)$$

where

$$\begin{aligned} \alpha_n(x) &= \mathbb{P}[Y_1^n = y_1^n, X_n = x], \\ \beta_n(x) &= \mathbb{P}[Y_{n+1}^N = y_{n+1}^N | X_n = x] \end{aligned}$$



are the so-called forward and backward variables, satisfying simple recurrence relations:

$$\begin{aligned}\alpha_{n+1}(x) &= \sum_{\tilde{x} \in A} \alpha_n(\tilde{x}) P_{\tilde{x},x} \Pi_{x,y_{n+1}}, \quad n = 1, \dots, N-1, \\ \text{with } \alpha_1(x) &= \mathbb{P}(X_1 = x) \Pi_{x,y_1}, \\ \beta_n(x) &= \sum_{\tilde{x} \in A} \beta_{n+1}(\tilde{x}) P_{x,\tilde{x}} \Pi_{\tilde{x},y_{n+1}}, \quad n = 1, \dots, N-1, \\ \text{with } \beta_N(x) &= 1.\end{aligned}$$

The key observation of [16] is that the probability distribution  $\mathbb{P}[X_n = \cdot | Y^N = y^N]$ , viewed as a column vector, can be expressed in terms of two-sided conditional probabilities  $\mathbb{Q}[Y_n = \cdot | Y^{N \setminus n} = y^{N \setminus n}]$ , with  $N \setminus n = \{1, \dots, N\} \setminus \{n\}$ , as follows

$$\begin{aligned}\mathbb{P}[X_n = \cdot | Y^N = y^N] \\ = \frac{\pi_{y_n} \odot \Pi^{-1} \mathbb{Q}[Y_n = \cdot | Y^{N \setminus n} = y^{N \setminus n}]}{\langle \pi_{y_n} \odot \Pi^{-1} \mathbb{Q}[Y_n = \cdot | Y^{N \setminus n} = y^{N \setminus n}], \mathbf{1} \rangle},\end{aligned}\quad (4.2)$$

where  $\Pi$  is the emission matrix, and  $\pi_{-1}, \pi_1$  are the columns of  $\Pi$ :

$$\begin{aligned}\Pi &= \begin{bmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{bmatrix}, \quad \pi_{-1} = \begin{bmatrix} 1-\epsilon \\ \epsilon \end{bmatrix}, \\ \pi_1 &= \begin{bmatrix} \epsilon \\ 1-\epsilon \end{bmatrix}, \quad \Pi^{-1} = \frac{1}{1-2\epsilon} \begin{bmatrix} 1-\epsilon & -\epsilon \\ -\epsilon & 1-\epsilon \end{bmatrix}\end{aligned}$$

and  $\odot$  is componentwise product of vectors of equal lengths,

$$u \odot v = (u_1 \cdot v_1, \dots, u_d \cdot v_d).$$

Expression (4.2) opens a possibility of constructing denoisers when parameters of the underlying Markov chains are unknown; we continue to assume that the

channel remains known. Indeed, two-sided conditional probabilities  $\mathbb{Q}[Y_n = \cdot | Y^{N \setminus n} = y^{N \setminus n}]$  could be estimated from the data. The Discrete Universal Denoiser algorithm (DUDE) [16] estimates conditional probabilities

$$\begin{aligned}\mathbb{Q}(Y_n = c | Y_{n-k_N}^{n-1} = a_{-k_N}^{-1}, Y_{n+1}^{n+k_N} = b_1^{k_N}) \\ = \frac{m(a_{-k_N}^{-1}, c, b_1^{k_N})}{\sum_{\tilde{c}} m(a_{-k_N}^{-1}, \tilde{c}, b_1^{k_N})}\end{aligned}\quad (4.3)$$

where  $m(a_{-k_N}^{-1}, c, b_1^{k_N})$  is the number of occurrences of the word  $a_{-k_N}^{-1} c b_1^{k_N}$  in the observed sequence  $y^N = (y_1, \dots, y_N)$ ; the length of right and left contexts is set to  $k_N = c \log N$ ,  $c > 0$ .

The DUDE has shown excellent performance in a number of test cases. In particular, in case of the binary memoryless channel and the symmetric Markov chain, considered in this paper, performance is comparable to the one of the backward-forward algorithm (4.1), which requires full knowledge of the source distribution, while DUDE is completely oblivious in that respect. In our opinion, the excellent performance of DUDE in this case is partially due to the fact that  $\mathbb{Q}$  is a Gibbs measure, admitting smooth two-sided conditional probabilities, which are well approximated by (4.3) and thus can be estimated from the data. It will be interesting to evaluate performance in cases when the output measure is not Gibbs.

Invention of DUDE sparked a great interest in two-sided approaches to information-theoretic problems. It turns out that despite the fact the efficient algorithms for estimation of one-sided models exist, the analogous two-sided problem is substantially more difficult. As alternatives to (4.3), other methods to estimate two-sided conditional probabilities have been suggested, e.g., [6, 11, 17]. For example, Yu and Verdú [17] proposed a **Backward-Forward Product** (BFP) model:

$$\tilde{\mathbb{Q}}(y_0 | y_{<0}, y_{>0}) \propto \tilde{\mathbb{Q}}(y_0 | y_{<0}) \tilde{\mathbb{Q}}(y_0 | y_{>0}),$$

and the one-sided conditional probabilities  $\tilde{\mathbb{Q}}(y_0 | y_{<0})$ ,  $\tilde{\mathbb{Q}}(y_0 | y_{>0})$  can be estimated using standard one-sided algorithms. Note, that in our model,

$$\begin{aligned}& \frac{\tilde{\mathbb{Q}}(y_0 | y_{<0}) \tilde{\mathbb{Q}}(y_0 | y_{>0})}{\tilde{\mathbb{Q}}(y_0 | y_{<0}) \tilde{\mathbb{Q}}(y_0 | y_{>0}) + \tilde{\mathbb{Q}}(\bar{y}_0 | y_{<0}) \tilde{\mathbb{Q}}(\bar{y}_0 | y_{>0})} \\ &= \frac{\cosh(Ky_0 + A(w_{-1})) \cosh(Ky_0 + A(w_1))}{\cosh(Ky_0 + A(w_{-1})) \cosh(Ky_0 + A(w_1)) + \cosh(-Ky_0 + A(w_{-1})) \cosh(-Ky_0 + A(w_1))}\end{aligned}$$

in general does not coincide with

$$\frac{\cosh(Ky_0 + A(w_{-1}) + A(w_1))}{\cosh(Ky_0 + A(w_{-1}) + A(w_1)) + \cosh(-Ky_0 + A(w_{-1}) + A(w_1))} = \mathbb{Q}(y_0 | y_{<0}, y_{>0}).$$

Nevertheless, the BFP model seems to perform extremely well [17].

Among other alternatives, let us mention the possibility to extend standard one-sided algorithms to produce algorithms for estimating two-sided conditional probabilities from data. This approach is investigated in a joint work with S. Berghout, where the denoising performance of the resulting Gibbsian models is evaluated. Gibbsian algorithm performs better than DUDE: bit error rates are given in the table below for noise level  $\epsilon = 0.2$  and various values of  $p$  (smaller rates are better).

$p$	Gibbs	DUDE
0.05	5.30 %	5.58 %
0.10	9.91 %	10.48 %
0.15	13.20 %	13.77 %
0.20	18.34 %	18.77 %

One could also try to estimate the Gibbsian potential directly, e.g., using the estimation procedure proposed in [5]. This method showed promising performance in experiments on language classification and authorship attribution. In conclusion, let us also mention that the direct two-sided Gibbs modeling of stochastic processes opens possibilities for applying semi-parametric statistical procedures, as opposed to the universal (parameter free) approach of DUDE.

#### Acknowledgments

Part of the work described in this paper has been completed during author's visit to the Institute of Mathematics for Industry, Kyushu University. The author is grateful for the hospitality during his stay and the support of the World Premier International Researcher Invitation Program.

Received: 27 August 2015 Revised: 29 November 2015

Accepted: 28 December 2015

Published online: 14 March 2016

#### References

- Behn, U, Zagrebnov, VA: One-dimensional Markovian-field Ising model: physical properties and characteristics of the discrete stochastic mapping. *J. Phys. A*. **21**(9), 2151–2165 (1988). ISSN:0305-4470, MR952930 (89j:82024)
- Berghout, S, Verbitskiy, E: On bi-directional modeling of information sources (2015)
- Bowen, R: Some systems with unique equilibrium states. *Math. Systems Theory*. **8**(3), 193–202. (1974/75). ISSN:0025-5661, MR0399413 (53 #3257)
- Ephraim, Y, Merhav, N: Hidden Markov processes, Special issue on Shannon theory: perspective, trends, and applications. *IEEE Trans. Inform. Theory*. **48**(6), 1518–1569 (2002). ISSN:0018-9448, MR1909472 (2003f:94024), doi:10.1109/TIT.2002.1003838
- Ermolaev, V, Verbitskiy, E: Thermodynamic Gibbs Formalism and Information Theory. In: *The Impact of Applications on Mathematics* (M. Wakayama et al, ed.), Mathematics for Industry, pp. 349–362. Springer Japan, (2014)
- Fernandez, F, Viola, A, Weinberger, MJ: Efficient Algorithms for Constructing Optimal Bi-directional Context Sets. *Data Compression Conference (DCC)*, 179–188 (2010). <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5453460>. doi:10.1109/DCC.2010.23
- Fernández, R, Ferrari, PA, Galves, A: Coupling renewal and perfect simulation of chains of infinite order, Ubatuba (2001)
- Hochwald, BM, Jelenković: State learning and mixing in entropy of hidden Markov processes and the Gilbert-Elliott channel. *IEEE Trans. Inform. Theory*. **45**(1), 128–138 (1999). ISSN:0018-9448, MR1677853 (99k:94028)
- Jacquet, P, Seroussi, G, Szpankowski, W: On the entropy of a hidden Markov process. *Theoret. Comput. Sci.* **395**(2–3), 203–219 (2008)
- Ordentlich, E, Weissman, T: Approximations for the entropy rate of a hidden Markov process. In: *Proceedings International Symposium on Information Theory 2005*, pp. 2198–2202, (2005). <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1523737>. doi:10.1109/ISIT.2005.1523737
- Ordentlich, E, Weinberger, MJ, Weissman, T: Multi-directional context sets with applications to universal denoising and compression. In: *ISIT 2005. Proceedings International Symposium on Information Theory*. pp. 1270–1274, (2005). <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1523546>. doi:10.1109/ISIT.2005.1523546
- Pollicott M: Computing entropy rates for hidden Markov processes. In: *Entropy of hidden Markov processes and connections to dynamical systems*, London Math. Soc. Lecture Note Ser, pp. 223–245, (2011). Cambridge Univ. Press, Cambridge MR2866670 (2012i:37010)
- Rabiner, LR: A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*. **77**, 257–286 (1989)
- Ruján, P: Calculation of the free energy of Ising systems by a recursion method. *Physica A: Stat. Theoretical Phys.* **91**(3–4), 549–562 (1978)
- Verbitskiy, EA: Thermodynamics of Hidden Markov Processes. In: Markus, B, Petersen, K, Weissman, T (eds.) *Papers from the Banff International Research Station Workshop on Entropy of Hidden Markov Processes and Connections to Dynamical Systems*, pp. 258–272. London Mathematical Society, Lecture Note Series, (2011). <http://dx.doi.org/10.1017/CBO9780511819407.010>
- Weissman, T, Ordentlich, E, Seroussi, G, Verdú, S, Weinberger, MJ: Universal discrete denoising: known channel. *IEEE Trans. Inform. Theory*. **51**(1), 5–28 (2005). ISSN: 0018-9448 MR2234569 (2008h:94036)
- Yu, J, Verdú, S: Schemes for bidirectional modeling of discrete stationary sources. *IEEE Trans. Inform. Theory*. **52**(11), 4789–4807 (2006). ISSN:0018-9448, MR2300356 (2007m:94144)
- Zuk, O, Domany, E, Kanter, I, Aizenman, M: From Finite-System Entropy to Entropy Rate for a Hidden Markov Process. *IEEE Sig. Proc. Letters*. **13**(9), 517–520 (2006)

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)